

## MEMORANDUM

Memorandum  
EDITED on February 26, 2002

### **Sample Design for 2002 NAEP**

Keith Rust, Westat  
August 2, 2001

The 2002 NAEP design is based on two fundamental components. The first is that there are four types of assessment sessions that must be administered at each of grades 4, 8, and 12. These have come to be referred to as A, B, C, and D. The second is that data from session type A are to be used for obtaining both national estimates and state-by-state public school estimates. The samples for each of these are to be integrated. This means that a public school student, from grade 4 or 8, in a participating state, who is selected for the A sample, will have his/her data used in producing both a state estimate and the national estimate for a given subject (reading or writing, depending upon the actual test booklet randomly assigned to the student). The B, C, and D samples are to represent the nation (50 states plus DC), but not individual states.

#### **1. Sample Type**

First consider the components of each sample type:

Sample Type A: This consists of the operational assessments for reading and writing for 2002. The data from students in sample type A will be used DIRECTLY to produce the national and state assessment results for 2002. There are two types of assessment booklets: reading and writing. In most cases, reading and writing books will alternate in the spiral, although there are more unique writing books than reading books. For grade 12 public schools, the A sample will be 64 percent writing books and 36 percent reading books.

Sample Type B: This consists of a variety of field test materials in reading and mathematics. More details are given in the ETS design memo, "2002 NAEP Field Activities of July 2001" (more detail is also given for the other sample types), available on the NAEP web site. A student will either take a reading book or a math book. At grade 12 there are only math books.

Sample Type C: This is a Reading assessment – exactly the same as the reading assessment conducted in 1998 (except that all SD/LEP students will be offered accommodations, instead of only doing this in half of the sampled schools). Data from this assessment will be used INDIRECTLY in the reporting of 2002 reading results for states and the nation, as these data will be used to equate the reading assessments from the A sessions to past assessments.

Sample Type D: This is a mathematics assessment, exactly the same as the mathematics assessment conducted in 2000 (except that all SD/LEP students will be offered accommodations, instead of only doing this in half of the sampled schools). Data from these assessments will be used to link Math results from 2000 to those from 2003.

## 2. Target Student Sample Sizes

Table 1 shows the student sample sizes needed for each sample type, grade, and school type. Note that for the grades 4 and 8 public A samples, the sizes are rather indeterminate. This is because of the requirement that these samples satisfy the needs of both the national estimates and those of the participating states. For purely national purposes, a sample of 32,000 students is required at each grade, proportionately allocated across the 50 states plus DC. For each participating state and jurisdiction, the sample size requirement is 5,000 per grade. As the largest state (California) has approximately one-eighth of the U.S. population, the state requirements supercede the national requirements in all participating states. Thus, the final sample size for each grade is 5,000 per participating state and jurisdiction, plus the proportional allocation of 32,000 for each nonparticipating state. At present there are between three and ten nonparticipating states. The three definite nonparticipants are very small (New Hampshire, South Dakota, and the BIA schools), but some of the uncertain ones are large (in particular, New York).

Table 1. Projected NAEP student sample sizes - 2002 assessment

Required numbers of assessed students					
<i>Grade 4</i>	Sample type				
	A	B	C	D	
Public	*	26,100	7,200	8,100	41,400
Catholic	6,000	1,450	400	450	8,300
Conservative Christian	1,500	435	120	135	2,190
Lutheran	1,500	145	40	45	1,730
Other Private	3,000	870	240	270	4,380
	12,000	29,000	8,000	9,000	58,000
<i>Grade 8</i>	Sample type				
	A	B	C	D	
Public	*	29,700	9,900	8,100	47,700
Catholic	6,000	1,650	550	450	8,650
Conservative Christian	1,500	495	165	135	2,295
Lutheran	1,500	165	55	45	1,765
Other Private	3,000	990	330	270	4,590
	12,000	33,000	11,000	9,000	65,000

Table 1. Projected NAEP student sample sizes - 2002 assessment (Continued)

<i>Grade 12</i>	Sample type				
	A	B	C	D	
Public	25,000	7,200	11,700	8,100	52,000
Catholic	6,000	400	650	450	7,500
Conservative Christian	1,500	120	195	135	1,950
Lutheran	1,500	40	65	45	1,650
Other Private	3,000	240	390	270	3,900
	37,000	8,000	13,000	9,000	67,000
<i>Total</i>	Sample type				
	A	B	C	D	
Public	25,000	63,000	28,800	24,300	141,100
Catholic	18,000	3,500	1,600	1,350	24,450
Conservative Christian	4,500	1,050	480	405	6,435
Lutheran	4,500	350	160	135	5,145
Other Private	9,000	2,100	960	810	12,870
	61,000	70,000	32,000	27,000	190,000

*\*State/National Combined Samples - Grade 4*

		All states	Half states	No states	
Per participating jurisdiction	5,000	280,000	140,000	0	
In national subsample	32,000	0	16,000	32,000	
Total		280,000	156,000	32,000	

*\*State/National Combined Samples - Grade 8*

		All states	Half states	No states	
Per participating jurisdiction	5,000	280,000	140,000	0	
In national subsample	32,000	0	16,000	32,000	
Total		280,000	156,000	32,000	
		All states	Half states	No states	
<i>Grand total</i>		750,000	502,000	254,000	

\* If all jurisdictions participate, no (additional) students needed for national subsample.

If half of the jurisdictions participate, then 16000 assessed students per grade needed, from nonparticipating states.

If no jurisdictions participate, then 32000 assessed students per grade needed in national sample.

A special spiral is needed for A sample Grade 12 Public (9/25 reading, 16/25 writing).

For all other cases the A sample is half reading and half writing.

### **3. School Samples**

To achieve the objectives of these samples, we are selecting four distinct types of school samples. We tend to try to name them by sample type, but in fact this can be quite misleading. So I am going to start calling them Alpha, Beta, Gamma and Delta.

#### **School Sample Alpha**

This consists of grade 4 and grade 8 public schools, selected to do sample type A for both national and state reporting. These samples are selected in two steps, at each grade:

1. First, a sample is selected for each state and jurisdiction, targeted to include 6,300 students in the final student sample. Exceptions are 1) if the jurisdiction does not have 6,300 students in the grade, all students are included; 2) if the state qualifies for and accepts the sparse state sample option, this target of 6,300 is reduced appropriately.
2. Then, within each of the 50 states plus DC, BIA, and DoDDS domestic schools, a subsample of the initial sample will be selected such that:
  - i. The student sample will be allocated in proportion to the population of the jurisdiction within the grade.
  - ii. The total student sample from the 53 entities will 32,000 students, after allowing for the rates of school and student nonresponse, school ineligibility, and student exclusion traditionally seen in national public school samples for reading and writing (such as the 1998 national assessment). In California this will mean that at least 80 percent of the initial sample schools will be included in this subsample, and perhaps all of them. There will be about 700 participating schools at grade 4, and 550 at grade 8 (remembering that these are subsamples of the state samples, and so do not actually add schools to the assessment except in nonparticipating states).

The subsample schools will be used in the following scenarios:

- i. If the state declines to participate in the state program, the schools in the subsample will still be part of the 2002 assessment. The other schools selected for the state will be dropped. So far there are two such states, South Dakota and New Hampshire. This is to be the case no matter how late the state decides to pull out of the assessment (such as when they find that they are not able to achieve a 70% participation rate).
- ii. It might turn out that NCES is not able to pay for Westat to administer the assessments for the state assessment program. This could be because the law is not changed to allow this or because insufficient funding is available. Then Westat will administer the assessments to the subsample of schools while the states will administer the rest. Note that in California Westat will still be assessing all (or almost all) the schools in the sample and a large proportion of the schools in other large states.
- iii. A further subsample of each of these subsamples will be used for the two special studies being conducted in 2002. A subsample (of as yet unspecified size) of the grade 4 national subsample will be identified to participate in the Oral Reading Study. A

subsample (of as yet unspecified size) of the grade 8 schools will be identified to participate in the grade 8 writing on line study.

Some additional features of sample Alpha are as follows:

- i. There is no oversampling of any subgroups, either of schools or within schools.
- ii. Small states will be allowed to reduce their sample size by reducing the sample size in their largest schools. Traditionally Delaware and Rhode Island have made use of this option; but if Westat is to administer all assessments, they may not wish to do so in 2002. Also, see point iv below.
- iii. There is no partitioning of samples into S2 and S3, etc, as in the past. There is also no assignment to subjects, either at the school level or within schools. Every session will be reading and writing.
- iv. At grade 4 a school with between 60 and 120 students may elect to have all students included in the sample or 60 students. See further discussion under Student Sampling. For design purposes we are assuming that no schools take this option, but in practice it is believed that the great majority of those offered it will take the option. If so, this will increase the grade 4 sample by an average of 20 percent with a high of 33 percent (Maryland). Obviously jurisdictions where there are fewer than 6,300 students will not experience any increase in sample size.
- v. Schools in sample Alpha may appear in sample Alpha for the other grade, sample Beta, and/or sample Gamma, depending upon what grades they have. We will be minimizing the overlap with the Beta sample at the same grade, otherwise it is just the luck of the draw. A school in DC or Delaware that has grades 4, 8, and 12 could conceivably end up in five samples – Alpha at grades 4 and 8 (this is certain to occur), Beta at grades 4 and 8, and Gamma at grade 12.

### **School Sample Beta**

This is also for public schools, grades 4 and 8. This sample will be stratified by jurisdiction, allocated in proportion to the population size. Overlap at the same grade with the Alpha sample will be minimized. The target for assessed students is 41,400 at grade 4, and 47,700 at grade 8. Thus the student sample sizes are somewhat larger than for the Alpha sample national subsample, but at the school level this will be mitigated by taking more students per school. Thus there are expected to be about 830 participating schools at grade 4, and 600 at grade 8.

At grade 4 each school will be assigned exactly one of the session types B, C, and D. These will be assigned in the ratio 29:8:9, so that 63 percent of schools will be doing the B session type. Note that this means that each “explicit” sampling stratum should have at least six schools selected to ensure that at least one school per stratum is assigned to each session type.

At grade 8, schools will be assigned from one to four session types depending upon their size. Session types will be assigned in the ratio 33:11:9. Thus no school will receive more than one session of either types C or D, but schools maybe assigned up to three sessions of B. A school could be assigned all three session types (two of B and one each of C and D). Again, it is highly desirable that each explicit sampling stratum have at least six schools assigned to it.

There will be no oversampling, either at the school or student level. There is no assignment to S2/S3. Within each grade, overlap with the Alpha samples will be minimized, but in places where all schools are included in the Alpha sample there will be some overlap, and such schools might therefore be asked to do all four session types if they are large enough.

### **School Sample Gamma**

This will be a single sample of grade 12 public schools, nationally representative of the 50 states plus DC, BIA, and DoDDS domestic schools. All sample types A, B, C, and D will be covered, although no single school will do more than three of these. The target number of assessed students is 52,000; although for operational reasons this design is to be modified slightly.

The sample is to oversample high minority schools (those with more than 15 percent black plus Hispanic enrollment) by a factor of two as has been done for national samples in the past. The sample is to minimize overlap with the sample of 400 grade 10 schools already selected for the NCES Educational Longitudinal Study (ELS, not conducted by Westat).

For student sampling, minority students are to be oversampled by a factor of two in low minority schools, at least for schools selected using the E-sample procedure, who identify minority students on the student file. There is no assignment to S2/S3. Stratification of the design has not yet been finalized. The sample size of participating schools is likely to be around 700.

This design calls for 25,000 students in the A samples, and 27,000 students in the B, C, and D samples combined. The design will be far more robust if the A sample constitutes exactly one half of the target. Therefore we have modified the target sample sizes as follows: A sample 27,300 (up from 25,000), B sample 7,800 (up from 7,200), C sample 11,700 (no change), D sample 7,800 (down from 8,100). This gives a total target of 54,600, rather than 52,000.

Schools will be assigned one to four “sessions” depending on their size. Session types A, B, C, and D will be assigned in the ratios 7:2:3:2. This means that a school with four sessions will have two sessions of A and one each of two of B, C, and D. Schools assigned three sessions will have one or two sessions of A, and one session each of two or one of B, C, and D. Schools assigned two sessions will have one of A and one of B, C, or D.

### **School Sample Delta**

This sample is for private schools at each of grades 4, 8, and 12. It will cover all session types A, B, C, and D. At grade 4 a school will be assigned a single session type. At grades 8 and 12, a school will be assigned no more than 2 session types (A and perhaps one other).

The primary stratification for each grade is as follows: Catholic, Conservative Christian, Lutheran, Other Affiliation, and Unknown Affiliation. For analysis, the schools with unknown affiliation will be placed in the appropriate category from among the other strata. The target student sample sizes by session type and by stratum are included in Table 1. For the Unknown Affiliation stratum, 25 schools will be selected at each grade. For this stratum the assignment of sessions has not been determined, but the great majority will be A sessions.

There is no oversampling within the primary strata identified above, and no assignment of S2/S3. The stratification within the primary strata varies, but in all cases is based on Census Division (there are nine of these), Type of Location (8 levels), minority enrollment, and size. The grade 12 sample will have minimum overlap with the ELS sample of 200 private schools (perhaps some Lutheran high schools will be in both studies).

Presumably a subsample of the grade 4 schools that are assigned to an A session will participate in the Oral Reading Study, while a subsample of the grade 8 schools assigned to do at least one A session will be assigned to do the Writing On Line Study. Sample sizes are not yet specified.

#### **4. Student Sampling**

##### **Sample Alpha, Grade 4 Schools**

1. All students, up to 64, will be selected.
2. If the school has between 65 and 120 students (inclusive) the school will be asked if it wishes to have all students selected, or a sample of 60. If the school elects to have all students selected, all students will be selected. If the school asks for a sample of 60 students, a systematic sample of 60 students will be selected with no oversampling.
3. If the school has more than 120 students, a systematic sample of 60 students will be selected with no oversampling.

##### **Sample Alpha, Grade 8 Schools**

1. All students, up to 64, will be selected.
2. If the school has more than 64 students, a systematic sample of 60 students will be selected with no oversampling.

##### **Sample Beta**

1. All students, up to 128, will be selected.
2. If the school has more than 128 students, a systematic sample of 120 students will be selected with no oversampling.
3. If the school has 30 or fewer students, it will be assigned one "session." If the school has 31 to 60 students, it will be assigned two sessions (but they may both be of type B). If the school has 61 to 90 students, it will be assigned three sessions (at least one, and usually two, will be type B). If the school has 91 students or more, it will be assigned four sessions, at least two and perhaps three, of which will be of type B.
4. Selected students will be assigned to sessions systematically.

### **Sample Gamma**

1. All students, up to 149, will be selected.
2. If the school has more than 149 students, but fewer than 273, a systematic sample of 136 students will be selected. If the school is low minority, then all unsampled black and Hispanic students will be added to the sample (at least in schools that use E-sampling; with luck not in “field sampled” schools pending NCES approval).
3. If the school has more than 272 students, a systematic sample of 136 students will be selected. If the school is low minority, then an additional sample of black and Hispanic students will be selected, such that the overall rate of selection for these students will be twice that of nonminority students (at least in schools that use E-sampling; with luck not in “field sampled” schools pending NCES approval).
4. If the school has 34 or fewer students, it will be assigned one “session.” If the school has 35 to 68 students, it will be assigned two sessions, one of which will be type A. If the school has 69 to 102 students, it will be assigned three sessions (at least one and often two will be type A). If the school has 91 students or more, it will be assigned four sessions, two of which will be of type A.
5. Selected students will be assigned to sessions systematically.

### **Sample Delta, Grade 4**

1. All students, up to 64, will be selected.
2. If the school has between 65 and 120 students (inclusive) the school will be asked if it wishes to have all students selected or a sample of 60. If the school elects to have all students selected, all students will be selected. If the school asks for a sample of 60 students, a systematic sample of 60 students will be selected with no oversampling.
3. If the school has more than 120 students, a systematic sample of 60 students will be selected with no oversampling.

### **Sample Delta, Grade 8**

1. All students, up to 64, will be selected.
2. If the school has more than 64 students, a systematic sample of 60 students will be selected with no oversampling.
3. If the school has 24 or fewer students, it will be assigned one “session.” If the school has more than 24 students, it will be assigned two sessions (at least one and perhaps both will be of type A).
4. Selected students will be assigned to sessions systematically.



### **Sample Delta, Grade 12**

1. All students, up to 64, will be selected.
2. If the school has more than 64 students, a systematic sample of 60 students will be selected, with no oversampling. It is possible that for some larger schools (most likely Lutheran), a sample of 120 students will be selected (with no oversampling).
3. If the school has 24 or fewer students, it will be assigned one “session.” If the school has more than 24 students, it will be assigned two sessions (at least one and perhaps both will be of type A).
4. Selected students will be assigned to sessions systematically.

## **5. Weighting Requirements**

### **The Alpha Samples**

These samples will have a single set of weights for each subject (reading and writing) applied to reflect school and student nonresponse, any trimming, and the random assignment to a reading or writing book. There will be a single replication scheme. It seems that a second set of weights will also be required. These will take the first set of weights, and poststratify them to independent population totals just as has been done for national assessments in the past (after incorporating the A session weights from Sample Delta).

No national/state linking weights are required, nor any different sets of reporting or “modular” weights. Weights will be required for the special studies, presumably poststratified after merging with the special study samples from the Delta samples.

### **The Beta Samples**

Session Type B does not require any weights. Types C and D must each be weighted, including poststratification, after merging with types C and D from the Delta samples.

### **The Gamma Sample**

Session Type A must be weighted, presumably including poststratification (after merging with the grade 12 Type A Delta sample).

Session Type B does not require weighting

Session Types C and D must each be weighted, including poststratification, after merging with types C and D from the Delta samples.

### **The Delta Samples**

Session Type A at grades 4 and 8 must be weighted, presumably including poststratification (after merging with the Alpha samples).

Session Type A at grade 12 must be weighted, including poststratification after merging with the Gamma sample type A.

Session Type B does not require weighting.

Session Types C and D must each be weighted, including poststratification, after merging with types C and D from the Beta (grades 4 and 8) and Gamma (grade 12) samples.

Note that although the B session types do not require weighting for any of the samples, they will probably need to be included up to a certain point as part of the checking of the other session types. The 2001 assessment provides a similar experience on which we can base our plans.